Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **MAR 2010**	2. REPORT TYPE	3. DATES COVERED **00-00-2010 to 00-00-2010**
4. TITLE AND SUBTITLE **Prediction Intervals for Future Performance**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **L-3 Communications at Air Force Research Laboratory,6030 S. kent St,Mesa,AZ,85212**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADA538937. Presented at the Proceedings of the Conference on Behavior Representation in Modeling and Simulation (19th), held in Charleston, South Carolina, 21 - 24 March 2010. Sponsored in part by AFRL, ARI, ARL, DARPA, & ONR. U.S. Government or Federal Rights License

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**2**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Prediction Intervals for Future Performance

Kelly M. Addis[a]
Michael Krusmark[a]
Tiffany S. Jastrzembski[b]
Kevin A. Gluck[b]
[a]L-3 Communications at Air Force Research Laboratory
[b]Air Force Research Laboratory
6030 S. Kent St.
Mesa, AZ 85212
480-988-6561
kelly.addis@mesa.afmc.af.mil, michael.krusmark@mesa.afmc.af.mil, tiffany.jastrzembski@mesa.afmc.af.mil, kevin.gluck@mesa.afmc.af.mil

Stuart Rodgers
AGS TechNet
10887 Miriam Lane
Dayton, OH 45458
937-903-0558
stu@agstechnet.com

Keywords: mathematical modeling, prediction intervals, performance, learning

1. The Predictive Performance Optimizer

Building on more than a century of research on human memory and performance, the Predictive Performance Optimizer (PPO) is a state-of-the-art cognitive tool to help decision-makers, instructors, and learners of all types to assess current performance and predict future performance by capturing the dynamics of human learning with basic cognitive science principles.

The PPO is a user-friendly software tool that can track performance over the course of a learner's training history for virtually any quantitative measure of performance. It generates performance predictions at specified future points in time, and allows users to visually and graphically assess and compare the impact of potential future training regimens. The PPO accomplishes this by utilizing a mathematical model for performance prediction (shown in Equation 1.1 below) inspired by the General Performance Equation (Anderson & Schunn, 2000).

$$Performance = S * St * N^{c} * T^{-d} \qquad \textit{(Equation 1.1).}$$

It comprises three main parts: the power law of learning (N^c), the power law of forgetting (T^{-d}), and a stability term (*St*) which captures the effects of practice and retention as they are spaced over time. The combination of these terms, along with a scaling factor (*S*), produces point predictions of future performance based on mathematical regularities in the learner's historical performance (for additional details, see Jastrzembski et al., 2009).

A major intended use of the PPO is to provide instructors and trainers with principled guidance concerning the readiness of their trainees. We will now frame PPO's practical relevance into a "just-in-time" training refresher scenario. Consider a training manager attempting to gauge how much training a warfighter must receive to ensure performance at or above a specified level of effectiveness before he may be deployed. The training manager may load the warfighter's unique training history into PPO to generate point predictions of future performance. The training manager can then assess whether adjustments must be made to the future training routine to meet the desired training goals.

Given the variability in human performance, generation of pure point predictions is insufficient in helping training managers make critical training decisions. One can imagine a scenario where a point prediction is at or very close to the effectiveness standard. Should the training regimen be deemed sufficient in that case? Is additional training heeded? Can we be confident that the performer will achieve that level of effectiveness at all? It is therefore necessary to provide training managers with scientifically-principled estimates of risk around the model's point predictions, to better guide decisions that have an impact in the real-world. We now turn to a discussion concerning how best to compute a prediction interval (PI) around the model's point predictions.

2. Prediction Interval Calculations

Rather than discrete point predictions, PIs provide a range of possible values of future performance, and thus offer the trainer a more complete picture of what outcomes future training regimens may possess. Identifying a method to compute a principled PI for our needs, however, is far from straightforward.

One issue we face is that we must balance two interacting effects: on one hand, human performance generally becomes less variable with increased practice; on the other hand, model predictions generally become less certain with longer lead times. A second issue is the limited existing data with which to validate the model's extended predictions. Related fields, such as economics, typically possess data spanning months or years, but few psychological studies examine data across time scales longer than a few days. A third problem is that there is little in the psychological literature which focuses on predicting performance at future times, and within that research, the incorporation of PIs on future performance is almost entirely absent. Thus, we lack sufficient exemplars to directly apply any one methodology to our situation, and have turned to other disciplines (e.g., econometrics and biostatistics), whose application to our situation is less straightforward, for guidance as a result. A final hurdle is maintaining the generality of the model. The model is intended to be used for predicting performance in a wide range of areas, and thus a large range of dependent variables. Accordingly, any methodology to compute PIs must not make mathematical assumptions that cannot be met with most measures of performance.

One method commonly used to generate PIs is the incorporation of a noise parameter into one or more parts of the model. In a computational model, this can be relatively straightforward, and the ACT-R framework has several extant noise parameters that can be utilized in a variety of situations. In our mathematical implementation, however, it is less obvious how to add in a noise parameter. As such, we are evaluating which terms in our mathematical model have a strong theoretical motivation to vary, and how these terms might interact with one another. For example, the learning rate and/or the forgetting rate might vary from one training session to the next based on fluctuations in the attentiveness of the warfighter or variability in the quality of the information in the briefing before the training session begins. However, one still has to determine the form and magnitude of the distribution from which to sample the noise. For this, we are investigating measures of variability in model fits to observed data that may be used to estimate the variability expected in future data.

The resulting PIs from this method, or any similar method, on predicted future performance provide an important tool for trainers and decision-makers by presenting a range of likely values for future performance. In our warfighter scenario, the training manager may decide to adopt a conservative criterion and use the worst likely performance shown by the PIs as a guide to impact future training needs. Such a criterion would ensure that the warfighter is most likely to actually perform at or above the desired level of effectiveness.

3. Summary

The question of how to properly calculate PIs for a mathematical model of performance and learning is a challenging one. The existing psychological literature offers little insight. We are, however, investigating a number of promising methods from related fields. Specifically, implementing noise in the model to generate variability is one of several promising possibilities. The development of an elegant method for calculating PIs for psychological performance data would hopefully encourage widespread use of such intervals as opposed to simple point predictions which inherently have unspecified certainty in their precise value. Our poster will present results from our ongoing explorations of these methods.

4. References

Anderson, J.R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science (Vol. 5)*, (pp. 1-34). Mahwah, NJ: Erlbaum.

Jastrzembski, T.S., Gluck, K.A., & Rodgers, S. (2009). Improving Military Readiness: A State-of-the-Art Cognitive Tool to Predict Performance and Optimize Training Effectiveness. In *Proceedings of I/ITSEC 2009.* Orlando, FL: NTSA.

Author Biographies

KELLY M. ADDIS is a Research Scientist with L-3 Communications at the Air Force Research Laboratory with a background in modeling of human memory and learning, and collaborator on the PPO project.

MICHAEL KRUSMARK is a Research Scientist with L-3 Communications at the Air Force Research Laboratory and collaborator on the PPO project.

TIFFANY S. JASTRZEMBSKI is a Research Psychologist at the Air Force Research Laboratory focused on developing the mathematical model in PPO to capture the dynamics of human memory.

KEVIN A. GLUCK is a Senior Research Psychologist at the Air Force Research Laboratory and enthusiastic collaborator on the PPO project.

STUART RODGERS is a computer scientist focused on implementing cognitive models of human performance and other adaptive, reactive, and autonomous software systems. He is Director at AGS TechNet, Dayton, OH.